National Library
of Medicine

PubMed

PubMed	Nucleotide	Protein	Genome	Structure	PopSet	Taxonomy	OMIM

Search PubMed ▼ for _____ Go Clear

Limits   Preview/Index   History   Clipboard

Display   Citation   ▼   Save   Text   Order   Add to Clipboard

Construction and analysis of a profile library characterizing groups of structurally known proteins.

Ogiwara A, Uchiyama I, Takagi T, Kanehisa M

Human Genome Center. University of Tokyo. Japan. ogi@nibb.ac.jp

A new sequence motif library StrProf was constructed characterizing the groups of related proteins in the PDB three-dimensional structure database. For a representative member of each protein family, which was identified by cross-referencing the PDB with the PIR superfamily classification. a group of related sequences was collected by the BLAST search against the nonredundant protein sequence database. For every group, the motifs were identified automatically according to the criteria of conservation and uniqueness of pentapeptide patterns and with a dual dynamic programming algorithm. In the StrProf library, motifs are represented by profile matrices rather than consensus patterns to allow more flexible search capabilities. Another dynamic programming algorithm was then developed to search this motif library. When the computationally derived StrProf was compared with PROSITE. which is a manually derived motif library in the best consensus pattern representation. the numbers of identified patterns were comparable. StrProf missed about one third of the PROSITE motifs. but there were also new motifs lacking in PROSITE. The new library was incorporated in SMART (Sequence Motif Analysis and Retrieval Tool). a computer tool designed to help search and annotate biologically important sites in an unknown protein sequence. The client program is available free of charge through the Internet.

MeSH Terms:
- Algorithms
- Amino Acid Sequence*
- Comparative Study
- Databases. Factual
- Molecular Sequence Data
- Peptide Library*
- Protein Structure. Tertiary*